

Article

E-Learning Courses Evaluation on the Basis of Trainees' Feedback on Open Questions Text Analysis

Dimitrios O. Tsimaras¹, Stylianos Mystakidis² , Athanasios Christopoulos³, Emmanouil Zoulias^{4,*} 
and Ioannis Hatzilygeroudis⁵ 

¹ Ministry of Education, Andrea Papandreou 37, 15180 Marousi, Greece

² School of Natural Sciences, University of Patras, 26504 Patras, Greece

³ Centre for Learning Analytics, Faculty of Natural Sciences, University of Turku, 20500 Turku, Finland

⁴ Health Informatics Laboratory, Faculty of Nursing, National and Kapodistrian University of Athens, 11527 Athens, Greece

⁵ Department of Computer Engineering & Informatics, University of Patras, 26504 Patras, Greece

* Correspondence: ezoulias@nurs.uoa.gr

Abstract: Life-long learning is a necessity associated with the requirements of the fourth industrial revolution. Although distance online education played a major role in the evolution of the modern education system, this share grew dramatically because of the COVID-19 pandemic outbreak and the social distancing measures that were imposed. However, the quick and extensive adoption of online learning tools also highlighted the multidimensional weaknesses of online education and the needs that arise when considering such practices. To this end, the ease of collecting digital data, as well as the overall evolution of data analytics, enables researchers, and by extension educators, to systematically evaluate the pros and cons of such systems. For instance, advanced data mining methods can be used to find potential areas of concern or to confirm elements of excellence. In this work, we used text analysis methods on data that have emerged from participants' feedback in online lifelong learning programmes for professional development. We analysed 1890 Greek text-based answers of participants to open evaluation questions using standard text analysis processes. We finally produced 7-gram tokens from the words in the texts, from which we constructed meaningful sentences and characterized them as positive or negative. We introduced a new metric, called acceptance grade, to quantitatively evaluate them as far as their positive or negative content for the online courses is concerned. We finally based our evaluation on the top 10 sentences of each category (positive, negative). Validation of the results via two external experts and data triangulation showed an accuracy of 80%.

Keywords: e-learning; online learning; educational data analysis; text analysis; educational programme evaluation



Citation: Tsimaras, D.O.; Mystakidis, S.; Christopoulos, A.; Zoulias, E.; Hatzilygeroudis, I. E-Learning Courses Evaluation on the Basis of Trainees' Feedback on Open Questions Text Analysis. *Educ. Sci.* **2022**, *12*, 633. <https://doi.org/10.3390/educsci12090633>

Academic Editors: Sandra Raquel Gonçalves Fernandes, Marta Abelha and Ana Teresa Ferreira-Oliveira

Received: 1 August 2022

Accepted: 15 September 2022

Published: 18 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The demands of the fourth industrial revolution [1] imply the need to reskill and upskill the workforce frequently and in mass scales to keep productivity, job satisfaction, and competitiveness at high levels [2]. This is also in line with the decision of the European Commission to introduce a key policy instrument in the European Skills Agenda for sustainable competitiveness, social fairness, and resilience [3]. A great contributor to these changes has been the rapid growth of distance online education programmes, which have been estimated to represent approximately 30% of the total education provision in Europe [4], a trend which has been naturally further accelerated because of the COVID-19 pandemic. Likewise, online professional development (OPD) is on the rise globally, both for flexibility and sanitary reasons [5].

Considering the above, assessing the effectiveness of such training programmes, especially from the learners' end, is of vital importance [6]. Indeed, for educational organisations

to maintain or even increase enrolments in online courses, it is essential to listen to their students' voices and accordingly meet their demands [7]. Moreover, after considering the educational paradigm shift that the recent pandemic outbreak brought, the need to identify robust and reliable approaches to evaluate online learning becomes more apparent than ever [8,9]. Finally, some researchers [10,11] also underline that e-learning and e-training programmes have, by definition, different structures and, thus, require alternative evaluation approaches.

To date, such evaluations have been traditionally performed with the aid of quantitative methods (e.g., surveys) based on instruments with closed questions (e.g., Likert scales) [12]. It is undeniable that surveys offer numerous advantages in educational research: (a) they are regarded as the most efficient method to gather the opinions of a large-scale sample, (b) they allow for statistical analysis which can provide considerably accurate generalisations, and (c) they are participant-friendly since participants are familiar with how to respond to them quickly and easily thanks to their multiple-choice Likert scale style. Therefore, such an evaluation approach can be particularly useful for a rough overview of the main course components. However, surveys also present a set of limitations. As the authors in [13,14] argue, student self-report data are often biased. This bias is due to the socially desirable answers that participants may offer in order to be viewed as good or favourable by the researcher. The proposed solution to eliminate the impact of this issue is the anonymisation of the surveys. Concerns are also raised about the reliability of the evaluation questionnaires, which greatly depends on the formulation of its items [15]. A proposed solution to overcome this drawback is to triangulate the data utilizing additional data collection methods [16]. To this end, the growing and widespread adaption of ICT in education enables educators and practitioners to gather diverse yet objective data that, if interpreted correctly, can greatly aid strategic decision making.

Sentiment analysis of a text is a rapidly growing scientific field with various applications. The human sensations of emotion, attitude, mood, affection, sentiment, opinion, and appeal all contribute to the basic categories of sentiment analysis of text [17]. The early applications of this approach are identified in the field of behavioural sciences [18], though, nowadays, relevant efforts can be identified in many other fields, including education, e.g., [19,20]. Some of the most notable benefits of integrating sentiment analysis techniques in education are as follows [21,22]: (a) it enables educators to understand their students' needs and preferences, (b) it helps to "break" the distance between the e-learners and the lecturers; this is particularly important after considering the lack of face-to-face communication (e.g., facial expressions) that the online learning context naturally restricts, and (c) it allows for the conduct of teaching and learning interventions aligned with the changes observed in students emotions. By gathering and analysing such information, institutional stakeholders can make informed, data-driven decisions at large concerning the design and development of their online platforms and courses.

Castro & Tumibay [23] summarise some of the most widely adopted methods to facilitate sentiment and opinion analysis using text mining methods, whereas Firmansyah et al. [24] connect them with the objectives that such techniques can serve when applied in the educational sector. However, a challenge that still governs such a process concerns the robustness of the integrated model, as the human-produced language is not lean, clean, and neat [25]. Likewise, students' feedback has often been found to be unstructured and inconclusive [26]. Indeed, every processing model which presumes stability, order, and consistency will break down when exposed to actual language use [27]. A model intended to accommodate new text should accept that every sentence is in order, as it derives, without preprocessing, re-editing, or normalisation, leaning on mechanisms keen on accepting new conventions, misspellings, nonstandard usage, and code switching. Models which rely on nontrivial knowledge-intensive preprocessing (such as part-of-speech tagging, syntactic chunking, named entity recognition, and language identification) or external resources (such as thesauri or ontologies) will always be brittle in view of real-world data [28].

In line with the efforts presented and discussed above, in this work, we build on the available models and introduce a methodology that can be utilised to improve the interpretation of participants' feedback as it emerges in the context of different OPD courses. It should be noted that the open-ended questions introduced for the needs of this work constitute an extension of the formal evaluation and are aimed at highlighting any unforeseen strengths or weaknesses of the delivered courses. The above outlines the main objective of the present work, which is to present a valid and reusable data-driven assessment method based on free comments, anonymous or eponymous, formulated in natural language. The novelty of this method is the use of text analysis for conducting a deeper and more meaningful e-learning evaluation based on open and free-flowing comments of participants. In consideration of the present findings, we hope that this effort will motivate trainers, e-learning managers, and administrators to consider similar adoptions in a more structured and systematic way as a means of reforming existing and new courses.

The structure of the paper is as follows. Section 2 presents our method for e-courses evaluation. Section 3 presents the experiment conducted and the produced results. In Section 4, validation issues are presented, and finally, Section 5 concludes the paper.

2. Evaluation Method

The method proposed in this paper concerns the evaluation of e-learning courses (or programmes) based on text analysis of the trainees' answers to open questions on an evaluation questionnaire. The questionnaire was distributed in the context of evaluation of the Center for Life-Long Learning (LLL) of the University of Patras (see next section for details). For the purposes of this study, data (sentences) were extracted from the answers to open-ended questions, containing users' impressions and suggestions for improvement, and stored in a comma-separated values (CSV) file. Before we present our method, we present the first two basic steps included in it.

2.1. Answer-Text Preprocessing

This first step is based on text analysis processes, as described below. We used RapidMiner to implement those processes. After the data CSV file is loaded in RapidMiner as an ExampleSet, the following processes are performed: "Tokenize" [29], "Transform Cases" [30], "Filter Stop Words by dictionary" [31], "Filter Tokens by length" [30], "Generate n grams" [32], (see Figure 1).

The first process is Tokenization. Tokenizing a document means splitting its text into individual elements or items, for example, words. We used the "Tokenize" function of the RapidMiner. There are a series of options to specify different splitting ways. We chose the option that splits text into single words. For example, the sentence "I already used Knowledge from the course in my Job", after tokenization, will produce the following words: "I", "already", "used", "Knowledge", "from", "the", "course", "in", "my", "Job".

The second process is called Cases Transformation, which aims to identify common words regardless of the typesetting style (lowercase, uppercase, mixed cases). In the present implementation, the cases of characters in the document were transformed to uppercase, using the respective operator. So, for example, the above tokens are transformed into "I", "ALREADY", "USED", "KNOWLEDGE", "FROM", "THE", "COURSE", "IN", "MY", "JOB". More generally, the words "Like", "LiKe", and "like" are transformed into the same uppercase word LIKE, which we keep once. Apart from not considering the typings of the same word as different tokens, another reason for using this process is the elimination of the accentuation of Greek words.

The third process, Filter Stop Words, is used to remove (Greek) stop words from the produced tokens. A comprehensive collection of 847 Greek stop words [33] was used as a dictionary for implementing this third process in RapidMiner. In the above sentence example, the tokens (words) that are removed are: "I", "ALREADY", "FROM", "THE", "IN", and "MY". So, what remains are: "USED", "KNOWLEDGE", "COURSE", "JOB".

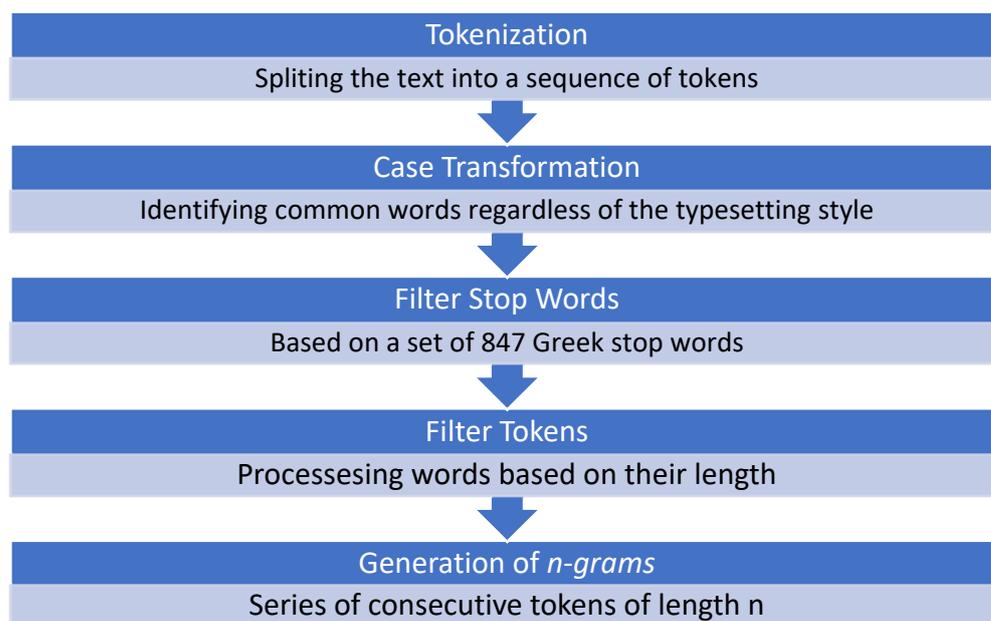


Figure 1. Answer-text preprocessing.

The fourth process, Filter Tokens, processes the words on the basis of their length (i.e., the number of characters each word contains). In the present implementation, three characters were set to be the minimum limit and 9999 characters the maximum. This was used to remove any very small size (less than three characters) left tokens (words), even after the third process application. None of the tokens from the above example are removed. In the Greek language, there are no two-character words that make any sense. All such words are included in the list of the Greek stop words we used and have already been removed in the previous step. This low limit has been set to exclude stop words possibly not included in the used list. The high limit was an arbitrary number to ensure that words of all lengths would be processed.

The final process concerns the generation of n grams. Generation of n grams from a vocabulary of tokens results in a series of combinations of tokens of length n . For example, from the above tokens, generation of 3 grams produces: “USED_KNOWLEDGE_COURSE”, “USED_KNOWLEDGE_JOB”, “USED_COURSE_JOB”, “KNOWLEDGE_COURSE_JOB”. That is, all possible combinations of three tokens are produced. We call the produced n grams phrases.

2.2. Phrase Evaluation

The above step applies to each open question answer-text by each trainee. The results of Rapid Miner are provided in the form of a worksheet, including several different word combinations in the form of n grams (phrases). For each phrase, the identity of each trainee who used it in their open answers is also stored.

In addition, the analysis includes information on the trainees’ answers to the closed question, “How would you describe your overall experience in the course?”. The answers to this question are important because each responder expresses their overall experience from the course. We call this question the golden question (GQ). It is a 10-scale question from grade 1 to 10 where grade 1 denotes a ‘not good experience’, and grade 10 is ‘an excellent experience’. As a result, the worksheet included the reply of each trainee to that question within the grade range (1 to 10).

From the resulting phrases (n grams) in the worksheet, those that were semantically related as answers to the golden question were selected by experts and split into two groups: emotionally positive and emotionally negative (see Section 4 for details). To evaluate those phrases, we introduce a new metric called the “acceptance grade”.

We consider the structure of data in the worksheet as depicted in Table 1. Each phrase (n gram) p_i is associated with one or more trainees t_j (those who used the phrase in their answers). Each trainee t_j is also associated with the grade g_j they gave as an answer to the golden question (GQ). Based on that, we define a new metric, called acceptance value (AV), as follows:

$$AV_{p_i} = \sum_j t_j g_j \text{ where } t_j \text{ is associated with } p_i. \tag{1}$$

Table 1. Text data structure.

No	Phrase (p_i)	Trainees (t_j)	GQ Grades (g_j)
1	phrase_1	trainee_2, trainee_3	grade_2, grade_3
2	phrase_2	trainee_1	grade_1
3	phrase_3	trainee_2, trainee_7, trainee_5	grade_2, grade_7, grade_5
4	phrase_4	trainee_5, trainee_16	grade_5, grade_16
5	phrase_5	trainee_7, trainee_20	grade_7, grade_20
...	...		
n	phrase_ n	trainee_ m , trainee_ k	grade_ m , grade_ k

For instance, the phrase “e-learning_made_easy_monitoring_progress” is one of the many different phrases that trainees use when answering the golden question. Let us suppose that the above phrase was included in answers to open questions of two (2) trainees who gave a 7 as an answer to GQ and one (1) trainee who gave a 4 to GQ. So, the phrase “e-learning made easy monitoring my progress” (let it be represented by p_x) has an acceptance value $AV_{p_x} = 2 \times 7 + 1 \times 4 = 18$. To get a normalised version, we define acceptance grade (AG) as

$$AG_{p_i} = AV_{p_i} / AV_{max}, \text{ where } AV_{max} \text{ is the maximum achieved acceptance value.}$$

In our example, dividing the AV_{p_x} score with the highest AV value (let say 150), we find that $AG_{p_x} = 18 / 150 = 0.12$.

Figure 2 illustrates the above process.

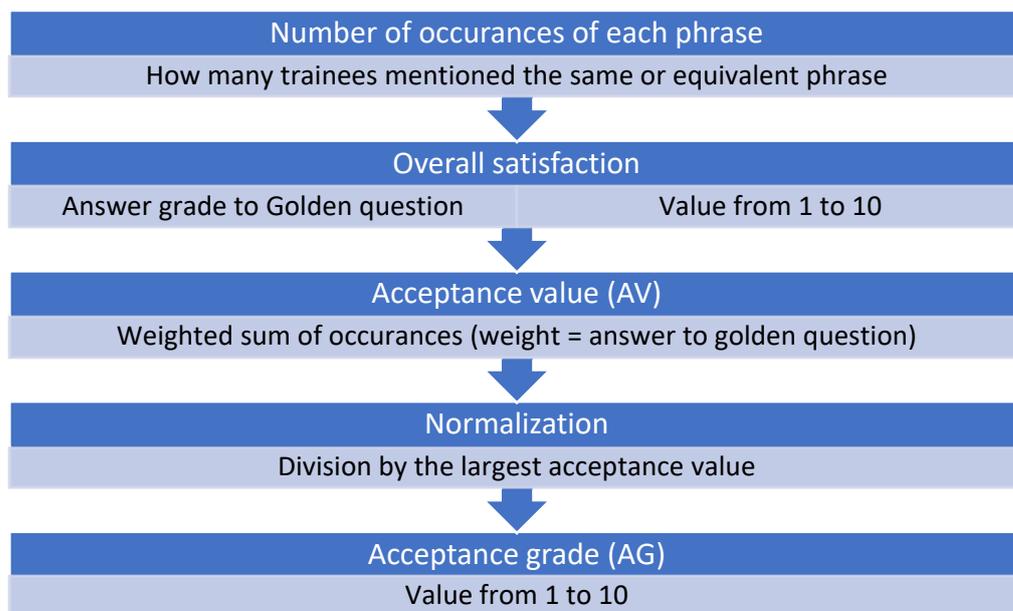


Figure 2. Phrase evaluation based on the calculation of acceptance grade.

2.3. Course Evaluation Method

Based on the above, we specify the evaluation method of a course on the basis of the answers of the trainees to the open questions addressed to them and the golden question, as illustrated in Figure 3.

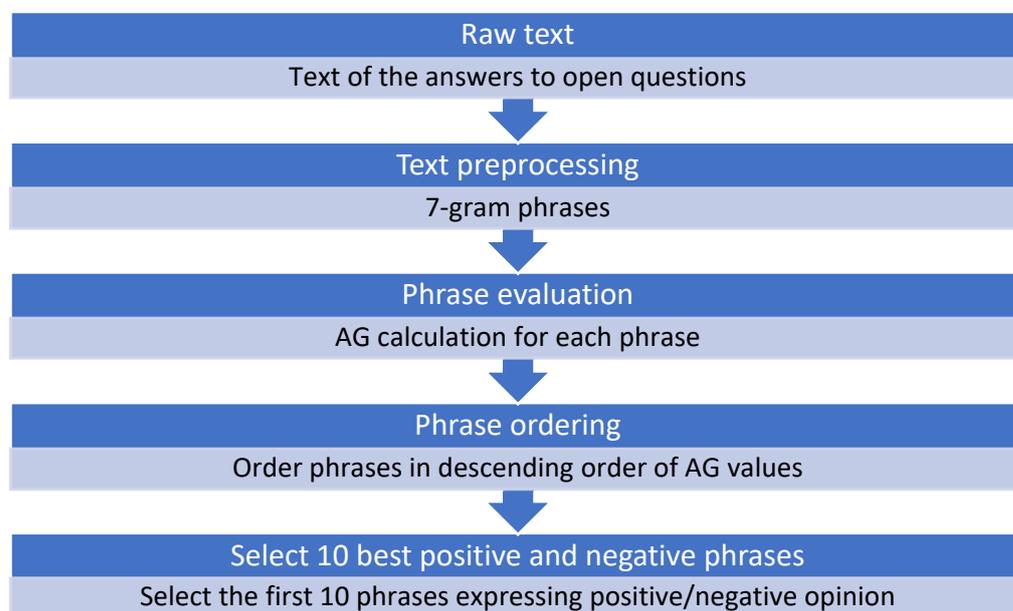


Figure 3. Course evaluation method.

In the first step, the raw text of the answers of the trainees to the open questions is provided. In the second step, the text preprocessing described in Section 3.1 is applied, resulting in the production of the 7-gram phrases. The decision to produce only 7 grams came from the observation that grams with $n < 7$ could not provide a basis for a meaningful phrase in the Greek language. This may change for another language. Afterwards, the AG of each 7-gram phrase is calculated. In the next step, phrases are ordered in descending mode with regard to their AG value. Finally, the first 10 phrases that indicate a positive or negative opinion about the course are identified and distributed in two groups, the positive group and the negative group (see next section).

3. Experiment and Results

3.1. Data Collection

This study was conducted in the context of evaluation of the Center for LLL (in Greek, ΚΕΔΙΒΙΜ-ΚΕΔΙΒΙΜ) of the University of Patras. ΚΕΔΙΒΙΜ authors and delivers multiple professional development e-learning courses dedicated to different topics and training needs [34].

The courses have been evaluated in line with the “Patras e-learning quality model” that was developed in the Center for LLL [35]. In line with the Content, Input, Process, Product (CIPP) model [36], the evaluation was performed on the grounds of the following axes: (a) the supportive framework of the course (infrastructure, content, support, organisation, and coordination), (b) the trainers (teaching performance), and (c) the course implementation (learning methods and results). For the evaluation, digital formative and summative questionnaires were utilised. The questionnaires were anonymous and consisted of 41 closed and 82 open items [35]. The total quantity of questions appears high because the questionnaires contain (a) grid questions, in other words, complex items with multiple aspects that learners rate, and (b) separate questions for each participating instructor. One question, for example, is formulated as “What is your satisfaction level with the following aspects of the course?” Although this is one question, it accounts for

15 items. In a similar fashion, approximately four items are dedicated per trainer, while each course is taught by 2 to 6 instructors. All participants responded to all questions as they were mandatory. A sample questionnaire is provided in the additional materials. The questionnaires featured 39 and 27 quality indicators (66 in total) on all aspects of the course design and delivery. The quality indicators were formulated either as an overall course component (e.g., assignment feedback) or as an individual trait (e.g., motivation provided by a specific trainer) to be rated on a scale from 1 to 5 (none, low, moderate, very good, excellent). Indicative examples of the open-ended questions are: "Comments on specific aspects of the action", "Specific comments on the trainers", "How would you describe your overall experience in the course?", "Summary of your overall experience & reviews", "Would you recommend this action to a colleague? What would you tell him/her?", "What would you suggest improving the program?".

In this work, we drew data from 27 online learning courses that follow a blended e-learning format in which both synchronous and asynchronous learning practices are combined. Data were collected from 20 evaluated vocational trainers' training courses in the field of Educational Sciences with 372 total participants. Each course featured at least four trainers and had a duration of 8 to 40 weeks. Most learners were female (70%). The main represented age groups were 25–34 years (54%), 35–44 (25%), and 45–49 (12%). Concerning their level of education, almost all held a higher education degree (97%), while 38% had an additional postgraduate degree. In total, 66% were employed, while 34% were seeking to re-enter the job market.

All courses were delivered using blended learning and had an overall completion rate of 85.48%. The formative questionnaire was distributed to the participants before the first half of each programme, whereas the summative questionnaire was administered after the completion of each programme. In each course, 4 trainers interacted with 16 trainees (on average) for a period of 8 to 16 weeks. In total, 378 responses were recorded, among them, 29% were male, and 71% were female. Most of the participants (98%) were holders of higher education degrees. As far as their professional identity is concerned, the participants were public and private sector employees, self-employed, and unemployed. The focus of the seminars included various teacher professional development training subjects.

3.2. Experiment

In total, 1890 documents, which were answers to open questions, including 4,838,056 words distributed across them, were processed for the analysis. We tried different n values ($n = 3-7$) for producing n -gram phrases. We calculated the acceptance grades (AGs) of all produced phrases. Although small n grams achieved greater values of AG, they could not be exploited for producing meaningful sentences.

The manual text analysis revealed that, although small n grams appear to have higher acceptance grades, they cannot be the basis for a meaningful statement. Further manual analysis revealed that 7 grams could be such a basis for the Greek language. As a result, we kept only the 7-word phrases, which reduced the n grams of interest to 5840 cases. Among the 5840 cases of 7 grams, a further context analysis was performed to identify phrases that have the same meaning. In case two phrases had the same meaning, the highest acceptance grade was kept and multiplied by the number of phrases with the same meaning. As a result, we had a new acceptance grade for the phrases that remained, and the cases were reduced to 5807.

Afterwards, we ordered all those phrases according to descending values of AG. So, phrases with the highest value of AG were first. Then, we began picking phrases, starting from the first one, putting them in one of two groups, the positive group and the negative group (also mentioned above). The criterion was whether a phrase indicated a positive or negative opinion about the course. Neutral phrases were bypassed. Next, the selected n -gram phrases were converted into English ones and then the English phrases were transformed into sentences by adding articles, propositions, and conjunctions.

3.3. Results

Table 2 summarises the produced sentences from the phrases in which trainees maintained a positive opinion over the deployed courses. As said, the phrases were translated by authors from the Greek 7 grams and transformed into English sentences using the words from the 7-gram phrases. For example, the Greek version of the 7-gram phrase corresponding to the second sentence in Table 2 is: “πλατφόρμα_ασύγχρονης_τηλεκπαίδευσης_έκανε_εύκολη_προσιτή_παρακολούθηση”; its English version is: “platform_asynchronous_elearning_made_easy_accessible_attendance”.

Table 2. Positive opinion sentences.

Sentence	Acceptance Grade
meets the demands of the modern multicultural society	10
A platform of asynchronous teletraining made attendance easy and accessible	8.89
A remarkable, very useful training programme that offered scientific knowledge	7.78
ability to apply in my work with the possibility of an apprenticeship	6.67
it offered me several new knowledge and skills	5.56
e-learning made it easy to monitor my progress	3.33
accessible learning while providing high-level education	3.33
the training programme is valuable, I have already used it in my work	2.22
I suggested this programme to colleagues	2.22
focusing more on practical and useful pedagogical methods	2.22

Interpreting the results in Table 2, we can say that participants agreed that the programmes meet the demands of the modern multicultural society (acceptance grade 10) and further underlined that the asynchronous education platform made monitoring easy and affordable (acceptance grade 8.89). Another important observation concerns the usefulness of the training programme in terms of scientific knowledge (acceptance grade 7.78) and practical application to their work (acceptance grade 6.67). Finally, the overly positive experience that participants had led them to also recommend it to colleagues and peers (acceptance grade 2.22).

Table 3 summarises the sentences in which trainees expressed a negative opinion of the courses. The most negative aspect seems to have been the extended length of the e-learning courses, which further highlighted the need to reduce its time (acceptance grade 10). Other concerns included the distraction participants experienced while attending the classes from their home environment (acceptance grade 7.92). In addition, participants mentioned that activities must be available every week (acceptance grade 5) and that trainers should acknowledge that the training programmes are for adults of basic education without advanced skills (acceptance grade 2.5).

Table 3. Negative opinion sentences.

Sentence	Acceptance Grade
boring lessons need to reduce lecture time	10
synchronous e-learning educators must be experts in platform use	8.33
time for questions and action tasks as in the last sessions	7.92
easy to be distracted at home in front of a screen	7.92
e-learning tiring when lasting many hours with ten-minute breaks	7.50
weekly activities must be available before the start of lessons	5
concerns adults of basic education so they can read	2.5
question and answer sessions should last longer	1.67
in e-learning, teachers should not talk continuously	1.67
somewhat pressing from the assignment’s deadlines point of view	1.67

4. Validation

The results that are depicted in Tables 2 and 3 represent the main opinion-charged statements expressed by participants that were produced by the proposed method. The validity of those results had to be assessed next.

For this purpose, two online learning external evaluation experts were nominated to validate the results. They were provided with the same raw user-evaluation data and were asked to come up with the main conclusions that emerged from the reflective feedback session data and the closed questions. The experts worked independently first and then synthesized their findings. Eventual discrepancies or disagreements were discussed to reach a consensus.

The reliability and validity of the research is also provided by data triangulation [37]. Data from two additional sources were obtained to compare the results of the proposed method. More specific, quantitative data from closed questions were also used to validate the subjective positive or negative claims of participants. Additionally, the final online synchronous meeting in each course was delivered as a reflective feedback session wherein the trainees had the opportunity to express their opinion, provide feedback, and locate strengths and weaknesses of the course [35]. Notes and observations from those sessions were used to assess the findings of the proposed text-analysis method. The evaluation experts worked at the level of each training course and provided a short evaluation report with the major findings pertaining to each course iteration. Subsequently, they compared findings from individual courses with the results from the proposed method.

The validation process concluded that the proposed method achieved 80% accuracy, capturing 8 of the 10 main emerging issues expressed in oral and written form. Specifically, on the positive side, several participants expressed their satisfaction with the course, verifying that it either met or exceeded their expectations regarding transferrable skills in the job market of a multicultural society. This is also verified quantitatively by the favourable overall evaluation in monitored quality metrics [35]. Other notable positive, opinion-charged statements pointed to participants' ability to monitor their progress and their readiness to recommend the course to other colleagues—a statement verified both in their feedback comments and the replies to a respective closed question.

On the other hand, in the context of academic freedom, in one early course iteration, a few academic educators did not participate in the suggested and quite demanding OPD on e-learning, relying on their prior experience with other systems. As a result, student experience was plagued by subpar engagement by trainers who relied excessively on a single teaching technique—the lecture. The second highest ranked issue of technical accessibility echoes technical challenges that these trainers faced with the technological affordances of the utilized platform. Other students asked for shorter online meetings and more opportunities to ask questions, even if this meant exceeding the prescribed duration of a meeting or scheduling additional online sessions. Topics and points that were missed can be attributed to the use of idiomatic language or spelling errors that were not identified successfully by the software.

5. Discussion and Conclusions

The systematic review of Choudhury & Pattnaik [38] examined the advantages and disadvantages of e-learning from the stakeholders' perspective. Amongst the key findings, the following major drawbacks are identified when it comes to assessing the educational potential of such programmes: (a) absence of effective evaluation methods and (b) difficulties in acquiring feedback from the learners. Although various methods and tools have been proposed for the design [39] and evaluation [40] of e-Learning programmes, an important aspect that is often not considered concerns the inclusion of "students' voice" in either of the aforementioned stages. In view of this shortcoming, the need to introduce advanced and dynamic approaches emerges.

In the present work, we gathered and classified online learners' feedback, as emerged from multiple courses, across different dimensions (e.g., attitude towards e-learning, course

design) and frequencies. The key findings revealed that the introduced programmes met their expectations both in terms of scientific knowledge advancement and practical application to their work. On the other hand, the most important shortcoming concerned the duration of the courses as well as the availability of the instructional activities on a weekly basis. Both of these findings are of particular importance to educational content designers and should be taken into account when preparing such courses [41]. Finally, regardless of the chosen course delivery method (physical or online), the difficulty level, as well as its escalation, are attributes that should be carefully analysed and evaluated prior to releasing a course. Indeed, a course cannot be too challenging, or people will experience it as frustrating. Yet, it should also not be too easy for learning to take place or, in other words, take place in students' zone of proximate development [42,43].

Notwithstanding the above, unstructured feedback in the form of a free-flowing written text has been historically regarded as impractical or counterproductive for the evaluation of training courses because of the difficulty and additional effort required to analyse each entry manually so as to detect possible themes and issues. However, open feedback without constraints allows learners to express themselves openly and to highlight what they think was most important or impactful within an OPD learning programme.

The methodology presented in this work can be integrated into any e-training scenario for a deeper evaluation, with minimal effort and preparation, as it is based on text analysis and classical calculus methods. In view of this, the vast evolution of speech-to-text software can greatly support similar efforts, especially when it comes to individuals with special needs or disabilities, as it enables them to submit their evaluation responses in the form of video or audio recordings, in cases where keyboard use is undesirable or tedious. Hence, the significance of this method will be further amplified in the future by the automatic transcription of verbal feedback comments, e.g., in video recordings, into written text.

The proposed text-analysis method can be used by practitioners and stakeholders to detect unforeseen problems or advantages and to reveal improvement ideas in the direction of transforming the evaluation procedure of OPD courses into iterative and mutually beneficial meaningful quality improvement instruments. This is not possible by using closed questions. One implication of the current study is the recommendation for wider use of open questions and the encouragement of participants' free-flowing feedback in e-learning course evaluation.

Further work can explore the proposed model with datasets emerging from participants with different geospatial (national level) or cultural backgrounds (international level). To this end, massive open online courses (MOOCs) can provide a great data stream source while also providing the opportunity to gather responses that emerge from professionals who are associated with different sectors (i.e., public and private).

In addition, a more technical vein of further work could be to automate the manual part of this method by semiautomating or automating the opinion mining process [44].

Author Contributions: Conceptualization, D.O.T. and E.Z.; Data curation, S.M.; Funding acquisition, D.O.T.; Investigation, S.M.; Methodology, D.O.T. and E.Z.; Project administration, D.O.T.; Resources, S.M.; Software, E.Z.; Supervision, D.O.T. and I.H.; Validation, E.Z.; Writing—original draft, S.M., A.C. and E.Z.; Writing—review & editing, S.M., A.C. and I.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Teo, T.; Unwin, S.; Scherer, R.; Gardiner, V. Initial Teacher Training for Twenty-First Century Skills in the Fourth Industrial Revolution (IR 4.0): A Scoping Review. *Comput. Educ.* **2021**, *170*, 104223. [\[CrossRef\]](#)
2. Mystakidis, S.; Papantzikos, G.; Stylios, C. Virtual Reality Escape Rooms for STEM Education in Industry 4.0: Greek Teachers Perspectives. In Proceedings of the 2021 6th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), Preveza, Greece, 24–26 September 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–5.
3. European Commission. *Communication on a European Skills Agenda for Sustainable Competitiveness, Social Fairness and Resilience*; European Commission: Brussels, Belgium, 2020.
4. Schneller, C.; Holmberg, C. *Distance Education in European Higher Education—The Offer*; Report 1 (of 3) of the IDEAL (Impact of Distance Education on Adult Learning) Project; UNESCO Institute for Lifelong Learning: Hamburg, Germany, 2014.
5. Bragg, L.; Walsh, C.; Heyeres, M. Successful Design and Delivery of Online Professional Development for Teachers: A Systematic Review of the Literature. *Comput. Educ.* **2021**, *166*, 104158. [\[CrossRef\]](#)
6. Powell, C.G.; Bodur, Y. Teachers' Perceptions of an Online Professional Development Experience: Implications for a Design and Implementation Framework. *Teach. Teach. Educ.* **2019**, *77*, 19–30. [\[CrossRef\]](#)
7. Castro, M.D.B.; Tumibay, G.M. A Literature Review: Efficacy of Online Learning Courses for Higher Education Institution Using Meta-Analysis. *Educ. Inf. Technol.* **2021**, *26*, 1367–1385. [\[CrossRef\]](#)
8. Firmansyah, R.; Putri, D.M.; Wicaksono, M.G.S.; Putri, S.F.; Widiyanto, A.A.; Palil, M.R. Educational Transformation: An Evaluation of Online Learning Due To COVID-19. *Int. J. Emerg. Technol. Learn. (ijET)* **2021**, *16*, 61–76. [\[CrossRef\]](#)
9. Umair, M.; Hakim, A.; Hussain, A.; Naseem, S. Sentiment Analysis of Students' Feedback before and after COVID-19 Pandemic. *Int. J. Emerg. Technol.* **2021**, *12*, 177–182.
10. McCutcheon, K.; Lohan, M.; Traynor, M.; Martin, D. A Systematic Review Evaluating the Impact of Online or Blended Learning vs. Face-to-Face Learning of Clinical Skills in Undergraduate Nurse Education. *J. Adv. Nurs.* **2015**, *71*, 255–270. [\[CrossRef\]](#)
11. Mystakidis, S.; Fragkaki, M.; Filippousis, G. Ready Teacher One: Virtual and Augmented Reality Online Professional Development for K-12 School Teachers. *Computers* **2021**, *10*, 134. [\[CrossRef\]](#)
12. Saadé, R.G.; He, X.; Kira, D. Exploring Dimensions to Online Learning. *Comput. Human Behav.* **2007**, *23*, 1721–1739. [\[CrossRef\]](#)
13. Hakkarainen, K.; Muukonen, H.; Lipponen, L.; Ilomaki, L.; Rahikainen, M.; Lehtinen, E. Teachers' Information and Communication Technology (ICT) Skills and Practices. *J. Technol. Teach. Educ.* **2001**, *9*, 181–197.
14. Rosenfeld, P.; Booth-Kewley, S.; Edwards, J.E.; Thomas, M.D. Responses on Computer Surveys: Impression Management, Social Desirability, and the Big Brother Syndrome. *Comput. Human Behav.* **1996**, *12*, 263–274. [\[CrossRef\]](#)
15. Means, B.; Toyama, Y.; Murphy, R.; Bakia, M.; Jones, K. *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies*; Association for Learning Technology: Oxfordshire, UK, 2010.
16. Cohen, L.; Manion, L.; Morrison, K. Planning Educational Research. In *Research Methods in Education*; Taylor and Francis: London, UK, 2011.
17. Karlgren, J.; Sahlgren, M.; Olsson, F.; Espinoza, F.; Hamfors, O. Usefulness of Sentiment Analysis. In Proceedings of the Advances in Information Retrieval, 34th European Conference on IR Research, ECIR 2012, Barcelona, Spain, 1–5 April 2012; Baeza-Yates, R., de Vries, A.P., Zaragoza, H., Cambazoglu, B.B., Murdock, V., Lempel, R., Silvestri, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 426–435.
18. Panksepp, J. Toward a General Psychobiological Theory of Emotions. *Behav. Brain Sci.* **1982**, *5*, 407–422. [\[CrossRef\]](#)
19. Lundqvist, K.; Liyanagunawardena, T.; Starkey, L. Evaluation of Student Feedback Within a MOOC Using Sentiment Analysis and Target Groups. *Int. Rev. Res. Open Distrib. Learn.* **2020**, *21*, 140–156. [\[CrossRef\]](#)
20. Bulusu, A.; Rao, K.R. Sentiment Analysis of Learner Reviews to Improve Efficacy of Massive Open Online Courses (MOOC's)—A Survey. In Proceedings of the 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 11 November 2021; pp. 933–941.
21. Berardinelli, N.; Gaber, M.; Haig, E. *SA-E: Sentiment Analysis for Education*; IOS Press: Amsterdam, The Netherlands, 2013; Volume 255.
22. Zhou, J.; Ye, J. Sentiment Analysis in Education Research: A Review of Journal Publications. *Interact. Learn. Environ.* **2020**, 1–13. [\[CrossRef\]](#)
23. Vidhya, K.A.; Aghila, G. Text Mining Process, Techniques and Tools: An Overview. *Int. J. Inf. Technol. Knowl. Manag.* **2010**, *2*, 613–622.
24. Ulfa, S.; Bringula, R.; Kurniawan, C.; Fadhli, M. Student Feedback on Online Learning by Using Sentiment Analysis: A Literature Review. In Proceedings of the 2020 6th International Conference on Education and Technology (ICET), Malang, Indonesia, 17 October 2020; pp. 53–58.
25. Sahlgren, M.; Karlgren, J. Terminology Mining in Social Media. In Proceedings of the 18th ACM Conference on Information and knowledge management, CIKM '09, Hong Kong, China, 2–6 November 2009; ACM Press: New York, NY, USA, 2009; p. 405.
26. Muralidharan, R.; Kanagasabapathy, T.; Vijai Ganesh, R.P. Analyzing ELearning Platform Reviews Using Sentimental Evaluation with SVM Classifier. *J. Phys. Conf. Ser.* **2021**, *1767*, 012012. [\[CrossRef\]](#)

27. Kastrati, Z.; Arifaj, B.; Lubishtani, A.; Gashi, F.; Nishliu, E. Aspect-Based Opinion Mining of Students' Reviews on Online Courses. In Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence, Tianjin, China, 23–26 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 510–514.
28. *Mining Text Data*; Aggarwal, C.C.; Zhai, C. (Eds.) Springer US: Boston, MA, USA, 2012; ISBN 978-1-4614-3222-7.
29. Grefenstette, G.; Tapanainen, P. What Is a Word, What Is a Sentence? Problems of Tokenisation. In Proceedings of the International Conference on Computational Lexicography, COMPLEX-94, Budapest, Hungary, 7–10 July 1994; pp. 79–87.
30. Lazarinis, F. Engineering and Utilizing a Stopword List in Greek Web Retrieval. *J. Am. Soc. Inf. Sci. Technol.* **2007**, *58*, 1645–1652. [[CrossRef](#)]
31. Baeza-Yates, R.A.; Ribeiro-Neto, B. *Modern Information Retrieval*; Addison-Wesley Longman Publishing Co., Inc.: Boston, MA, USA, 1999; ISBN 020139829X.
32. Zamora, E.M.; Pollock, J.J.; Zamora, A. The Use of Trigram Analysis for Spelling Error Detection. *Inf. Process. Manag.* **1981**, *17*, 305–316. [[CrossRef](#)]
33. Diaz, G. Greek Stopwords Collection. Available online: <https://github.com/stopwords-iso/stopwords-el> (accessed on 15 May 2022).
34. Mystakidis, S.; Berki, E.; Valtanen, J.-P.; Amanatides, E. Towards a Blended Strategy for Quality Distance Education Life-Long Learning Courses—The Patras Model. In Proceedings of the 17th European Conference on e-Learning (ECEL), Athens, Greece, 1–2 November 2018; pp. 408–416.
35. Mystakidis, S.; Berki, E.; Valtanen, J.-P. The Patras Blended Strategy Model for Deep and Meaningful Learning in Quality Life-Long Distance Education. *Electron. J. e-Learn.* **2019**, *17*, 66–78. [[CrossRef](#)]
36. Stufflebeam, D.L.; Zhang, G. *The CIPP Evaluation Model: How to Evaluate for Improvement and Accountability*; Guilford Publications: New York, NY, USA, 2017; ISBN 9781462529247.
37. Thurmond, V.A. The Point of Triangulation. *J. Nurs. Scholarsh.* **2001**, *33*, 253–258. [[CrossRef](#)] [[PubMed](#)]
38. Choudhury, S.; Pattnaik, S. Emerging Themes in E-Learning: A Review from the Stakeholders' Perspective. *Comput. Educ.* **2019**, *144*, 103657. [[CrossRef](#)]
39. Affouneh, S.; Salha, S.; Khlaif, Z.N. Designing Quality E-Learning Environments for Emergency Remote Teaching in Coronavirus Crisis. *Interdiscip. J. Virtual Learn. Med. Sci.* **2020**, *11*, 135–137. [[CrossRef](#)]
40. Lu, D.-N.; Le, H.-Q.; Vu, T.-H. The Factors Affecting Acceptance of E-Learning: A Machine Learning Algorithm Approach. *Educ. Sci.* **2020**, *10*, 270. [[CrossRef](#)]
41. Al-Fraihat, D.; Joy, M.; Masa'deh, R.; Sinclair, J. Evaluating E-Learning Systems Success: An Empirical Study. *Comput. Human Behav.* **2020**, *102*, 67–86. [[CrossRef](#)]
42. Ferguson, C.; van den Broek, E.L.; van Oostendorp, H. AI-Induced Guidance: Preserving the Optimal Zone of Proximal Development. *Comput. Educ. Artif. Intell.* **2022**, *3*, 100089. [[CrossRef](#)]
43. Sampayo-Vargas, S.; Cope, C.J.; He, Z.; Byrne, G.J. The Effectiveness of Adaptive Difficulty Adjustments on Students' Motivation and Learning in an Educational Computer Game. *Comput. Educ.* **2013**, *69*, 452–462. [[CrossRef](#)]
44. Misuraca, M.; Scepi, G.; Spano, M. Using Opinion Mining as an Educational Analytic: An Integrated Strategy for the Analysis of Students' Feedback. *Stud. Educ. Eval.* **2021**, *68*, 100979. [[CrossRef](#)]